ANESTIS ANTONIADIS

Taux de convergence pour des modèles de régression à inf-convolution splines et applications

Statistique et analyse des données, tome 16, nº 2 (1991), p. 1-12.
<http://www.numdam.org/item?id=SAD_1991__16_2_1_0>

Statistique et Analyse des Données
1991 - Vol. 16 n° 2 p. 1 - 12

TAUX DE CONVERGENCE POUR DES MODELES DE REGRESSION A INF-CONVOLUTION SPLINES ET APPLICATIONS

Anestis ANTONIADIS

Laboratoire IMAG-LMC
Université Joseph Fourier, B.P. 53X
38042 GRENOBLE
E-mail address : antonia@imag.fr

Résumé.

Un des domaines de recherche extrêmement vivant aujourd'hui en statistique mathématique est celui de la régression semi-paramétrique. Une variété de modèles semi-paramétriques est la classe des modèles de régression partiellement linéaires pour lesquels la fonction de régression contient un terme paramétrique linéaire et un terme additif non paramétrique dépendant d'une ou plusieurs co-variables. Dans ce travail nous nous intéressons à l'estimation efficace des coefficients de régression pour ce type de modèles sans pour cela sous-lisser l'estimateur inf-convolution spline de la composante non paramétrique.

Key words and phrases. Régression non paramétrique, inf-convolution, splines partielles, taux de convergence...

1991 Mathematics Subject Classification :
Primaire 62G99, 62J99 ; Secondaire 62E20.

Manuscrit reçu sous forme définitive le 22 janvier 1992.

1. INTRODUCTION.

Le modèle de régression que nous étudions dans cet article peut être décrit mathématiquement par les équations et les définitions suivantes. Supposons que les quantités observées $(y_1, y_2,...,y_n)$ sont une réalisation du vecteur aléatoire $(Y_1, Y_2,..., Y_n)$ dont chacune des composantes peut s'écrire sous la forme

$$Y_i = {}^t\beta x_i + f(t_i) + \varepsilon_i\ ,\ \ i = 1,...,n. \qquad (1.1)$$

où les x_i sont des vecteurs connus de dimension p, β est le vecteur p-dimensionnel inconnu des coefficients de régression, f une fonction régulière d'une variable réelle t et où les termes d'erreurs ε_i sont des variables aléatoires réelles indépendantes, centrées et de même variance σ^2 finie. L'objectif fixé est d'estimer efficacement les paramètres inconnus β et f du modèle de régression défini par l'équation (1.1). Nous serons conduits pour cela à faire des hypothèses supplémentaires sur le comportement régulier de f.

Signalons, à titre d'exemple, une situation pour laquelle la modélisation que nous présentons peut être appliquée. Considérons un échantillon de pièces ayant fonctionné dans un environnement donné et présentant un certain taux d'usure variable. Cet échantillon est partagé en deux groupes soumis à des traitements distincts. Notons $Y_1,...,Y_{n_1}$ les taux d'usure des pièces soumises au traitement de type 1 après avoir fonctionné un certain temps T et $Y_{n_1+1},...,Y_n$ les taux d'usure du lot de pièces restantes. Il est raisonnable alors d'analyser ces données en posant

$$Y_i = \beta x_i + g(\tau_i) + \varepsilon_i\ \ ,\ \ i = 1,...,n \qquad (1.2)$$

où $x_i = 1$ ou $x_i = 0$ selon le type de traitement subit par la pièce numéro i, et où $g(\tau_i)$ représente le taux d'usure d'une pièce à l'instant T lorsque elle présente un taux d'usure τ_i à l'instant initial. On pourrait bien sûr citer d'autres exemples réalistes et convaincants qui rentrent dans le cadre du modèle semi-linéaire de régression postulé en (1.1), mais cela ne présente que peu d'intérêt pour la suite.

Sans perdre de généralité, nous supposerons que les instants d'observations t_i varient dans l'intervalle [0,1]. En outre nous exigeons dans la suite un comportement régulier pour la fonction inconnue f. Plus précisément, f est supposée appartenir à l'espace de Sobolev d'ordre m (m entier positif)

$$\mathcal{H}^m[0,1] = \{f : [0,1] \to \mathbb{R} \; ; f^{(j)}, j = 0,...,m\text{-}1 \text{ absolument continues et } \int_0^1 [f^{(m)}(t)]^2 dt < \infty\} \; .$$

L'estimation des paramètres β et f est obtenue par minimisation selon $(\beta,g) \in \mathbb{R}^p \times \mathcal{H}^m[0,1]$ de la fonction de perte

$$\frac{1}{n}\sum_{i=1}^{n} (Y_i - {}^t\beta x_i - g(t_i))^2 + \lambda \int_0^1 [g^{(m)}(t)]^2 \, dt \tag{1.3}$$

Dans l'équation (1.3), le paramètre de lissage λ contrôle le compromis entre la fidélité aux données mesurée par le premier terme de la somme et le caractère lisse de la solution mesuré par l'intégrale $\int_0^1 [g^{(m)}(t)]^2 \, dt$. Pour s'apercevoir de l'importance de ce terme de pénalisation considérons le cas $m = 1$. Si $\lambda = 0$, alors l'expression (1.3) est minimisée en prenant $\hat{\beta} = 0$, $\hat{f}(t_i) = Y_i$ et f linéaire sur $[t_i,t_{i+1}[$. Un tel estimateur de β indépendant des observations faites est généralement indésirable. En prenant maintenant $m = 2$ et en considérant le cas où $\lambda \to \infty$, on voit que la solution optimale est forcée d'être linéaire et on tend alors vers la droite des moindres carrés ajustant les données.

L'estimateur obtenu par minimisation de (1.3), lorsqu'il existe, est connu sous le nom *inf-convolution spline* ou encore *spline partielle* . De tels estimateurs ont été considérés et étudiés par plusieurs auteurs, par exemple, Laurent et Girard [3], Rice et al [2], Wahba ([10]), Heckman [5], Speckman [8] et Eubank [4], pour n'en citer que quelques uns. Notons seulement ici, que la solution $\hat{f}$ du problème de minimisation posé est une fonction spline puisque, pour toute valeur fixée du paramètre linéaire β, l'objectif est de minimiser

$$\frac{1}{n}\sum_{i=1}^{n} (Z_i - g(t_i))^2 + \lambda \int_0^1 [g^{(m)}(t)]^2 \, dt \tag{1.4}$$

avec $Z_i = Y_i - {}^t\beta x_i$, et que la solution de ce problème, lorsque $n > m$, existe et est une fonction spline naturelle d'ordre 2m (cf. Laurent [6]). A ce titre et parce que nous en aurons besoin ultérieurement, rappelons qu'une fonction spline naturelle d'ordre 2m sur [0,1] basée sur les noeuds $t_1, t_2,...t_n$ est une fonction s(t) telle que :

- s est un polynôme de degré 2m-1 sur chacun des intervalles $[t_i,t_{i+1}[$, $i = 1,...,n-1$.
- s est un polynôme de degré m-1 sur chacun des intervalles $[0,t_1[$ e t $[t_n,1]$.
- $s \in C^{2m-2}(\mathbb{R})$

Il est connu que l'espace S_n^m des fonctions splines naturelles d'ordre 2m en les noeuds $t_1, t_2, \ldots, t_n$ est un espace vectoriel de dimension n. Ceci permet de voir que les estimateurs de β et f dans (1.3) peuvent être explicitement calculés en résolvant un système d'équations linéaires.

Notre propos dans les paragraphes suivants est d'étudier le comportement asymptotique de ce type d'estimateurs lorsque $n \to \infty$. Peu de résultats existent à ce propos dans la littérature. Rice ([7]) a montré que le biais de $\hat{\beta}$ peut asymptotiquement dominer sa variance lorsqu'on désire estimer la partie non paramétrique avec le taux de lissage optimal. Afin d'obtenir un estimateur consistant, on est donc conduit à sous-lisser la composante non paramétrique du modèle. Speckman ([9]) réussit à obtenir un estimateur convenable de β sans sous-lisser la partie non paramétrique, en utilisant la méthode du noyau, mais remarque que sa méthode ne peut être étendue au cas des splines partielles que dans le cas où f est supposée être périodique. Nous démontrerons que le résultat de Speckman s'étend dans le cadre des splines partielles sans hypothèses supplémentaires sur f. Dans le paragraphe suivant nous rappelons d'abord quelques résultats sur les splines de lissage classique $(\beta = 0)$, pour traiter ensuite du cas général.

2. LES SPLINES DE LISSAGE $(\beta = 0)$.

Nous présentons dans ce paragraphe une description résumée des résultats asymptotiques de l'estimateur obtenu après minimisation de la fonction de perte de l'introduction, lorsqu'on impose a priori que $\beta = 0$. Signalons, que le lecteur pourra trouver dans l'ouvrage récent et extrêmement complet de Eubank ([4]) des compléments à l'exposé qui est fait dans ce paragraphe. Sans perdre de généralité nous supposerons que les instants d'observation t_i sont distribués sur [0,1] de manière régulière, selon une densité de probabilité ω, c'est-à-dire

$$\frac{1}{n} \sum_{i=1}^{n} c(t_i) \to \int_0^1 c(t)\, \omega(t)dt$$

quand $n \to \infty$, pour toute fonction c continue sur [0,1].

Nous avons vu dans l'introduction que, pour $n > m$, l'estimateur spline $\hat{g}_\lambda(t)$ existe et est unique élément de l'espace des splines naturelles S_n^m. Il est donc de la forme

$$\hat{g}_\lambda(t) = \sum_{i=1}^{n} \hat{c}_{\lambda,i}\phi_i(t)$$

où $\{\phi_i\}_{i=1,\dots,n}$ est une base de S_n^m. Le vecteur $\hat{c}_\lambda$ des coefficients est solution du système linéaire

$$({}^t XX + n\lambda\Omega)\,\hat{c} = {}^t XY \tag{2.1}$$

où $X = (\phi_j(t_i))_{i,j=1,\dots,n}$ et $\Omega_{i,j} = \int_0^1 \phi_i^{(m)}(t)\,\phi_j^{(m)}(t)\,dt$. On obtient ainsi pour la solution de l'équation (2.1)

$$\hat{c}_\lambda = S(\lambda)Y \text{ avec } S(\lambda) = X({}^t XX + n\lambda\Omega)^{-1}\,{}^t X \tag{2.2}$$

Le critère adopté pour juger des bonnes qualités de l'estimateur spline obtenu est le risque intégré moyen (RIM)

$$R_n(\lambda) = \int_0^1 E\,(g(t) - \hat{g}_\lambda(t))^2\,\omega(t)\,dt \tag{2.3}$$

En décomposant le risque intégré moyen en somme d'un terme associé au biais et un terme de variance on obtient

$$R_n(\lambda) = B_n^2(\lambda) + V_n(\lambda),$$

où

$$B_n^2(\lambda) = \int_0^1 (g(t) - E\,(\hat{g}_\lambda(t)))^2\,\omega(t)\,dt$$

et

$$V_n(\lambda) = \int_0^1 \mathrm{var}(\hat{g}_\lambda(t))\,\omega(t)\,dt\ .$$

Pour $m \geq 2$, Speckman a montré dans [9], que

$$B_n^2(\lambda) \leq \lambda \int_0^1 (g^{(m)}(t))^2\,dt(1 + o(1))$$

quand $n \to \infty$, $\lambda \to 0$ et $n^2\lambda \to \infty$. On en déduit alors le taux de convergence $B_n^2(\lambda) = \mathcal{O}(\lambda)$. Pour le contrôle de la variance, le choix d'une base particulière de l'espace des splines naturelles permet de contrôler les valeurs propres de l'opérateur Ω et on obtient (voir Speckman [9]) $V_n(\lambda) = \mathcal{O}(n^{-1}\lambda^{-1/2m})$. Par conséquent, en prenant $\lambda = \mathcal{O}(n^{-\frac{2m}{2m+1}})$, le risque intégré moyen décroît vers 0 à la vitesse $\mathcal{O}(n^{-\frac{2m}{2m+1}})$.

La borne supérieure utilisée pour le carré du biais de l'estimateur $\hat{g}_\lambda$ peut souvent être améliorée. Ceci dépend bien sûr du comportement de la vraie fonction g aux bords du domaine $[0,1]$ et de son vrai degré de régularité. Rappelons ici que $\hat{g}_\lambda$ est une fonction spline naturelle d'ordre $2m$ et qu'en tant que telle elle satisfait aux conditions aux bords

$$\hat{g}_\lambda^{(j)}(0) = \hat{g}_\lambda^{(j)}(1) = 0, \quad j = m,\ldots,2m-1. \tag{2.4}$$

On peut donc espérer que lorsque la vraie fonction g est dans l'espace $\mathcal{H}^{2m}[0,1]$ et qu'elle satisfait aux conditions (2.4), le biais décroît plus rapidement. Pour le cas d'une densité uniforme $\omega(t) = 1$, on peut montrer (cf. [9]) que

$$B_n^2(\lambda) \sim \lambda^2 \int_0^1 (g^{(2m)}(t))^2\, dt$$

et dans ce cas le λ optimal est de l'ordre $\mathcal{O}(n^{-\frac{2m}{4m+1}})$ et le RIM moyen optimal est alors

$$R_n(\lambda) \sim \mathcal{O}(n^{-\frac{4m}{4m+1}})$$

Cette relation du biais des splines de lissage avec les conditions aux bords de g est l'analogue des problèmes de biais aux bords associés aux estimateurs du noyau. Pour les méthodes du noyau, des corrections locales aux bords de l'estimateur permettent souvent des meilleurs taux de convergence. Nous nous proposons dans le paragraphe suivant de montrer que ceci est également possible avec les splines de lissage et les splines partielles.

3. CORRECTION DU BIAIS.

Nous allons d'abord nous placer dans le cadre de la statistique inférentielle classique. Le résultat obtenu guidera ensuite notre choix de la méthode d'estimation adoptée pour le cas des splines de lissage et des splines inf-convolution.

Soit

$$(\Omega, \mathcal{A}, \{P_\theta ; \theta \in \mathbb{R}^p\})$$

une structure statistique paramétrique donnée, et soient $\hat{\theta}$ et $\bar{\theta}$ deux estimateurs distincts du paramètre p-dimensionnel inconnu θ. Imaginons que l'estimateur $\hat{\theta}$ est sans biais, mais présente une variance importante, alors que le second estimateur est biaisé, mais de variance raisonnable. Notre objectif est de construire un nouvel estimateur de θ qui présente les avantages des deux autres. Pour cela la qualité de l'estimateur sera mesurée par le risque quadratique moyen

$$pR(\tilde{\theta}) = E\{\|\tilde{\theta}-\theta\|^2\} = \|E\{\tilde{\theta}\}-\theta\|^2 + \text{trace } V(\tilde{\theta}) .$$

Si X est une matrice réelle, on notera P_X le projecteur orthogonal sur l'espace vectoriel engendré par les colonnes de X. Avec ces notations, on a :

Proposition 3.1.

Soit K une matrice réelle connue d'ordre $p \times k$ est supposons que $E\{\hat{\theta}\} = \theta$ et que $E\{\bar{\theta}\} = \theta + K\gamma + \xi$ où γ et ξ sont des paramètres inconnus. Alors

$$\tilde{\theta} = \bar{\theta} + P_K(\hat{\theta}-\bar{\theta}) \tag{3.1}$$

est un estimateur tel que

$$\|E\{\tilde{\theta}\} - \theta\|^2 \leq \|\xi\|^2$$

et

$$\text{trace } V(\tilde{\theta}) \leq \text{trace } V(\bar{\theta}) + \text{rang}(K)\mu_{max}(V(\hat{\theta}))$$

où μ_{max} désigne la plus grande valeur propre de X.

Preuve.

Par un simple calcul algébrique on a

$$\|E\{\tilde{\theta}\} - \theta\|^2 = \|(I - P_K)\,[E\{\bar{\theta}\} - \theta]\|^2 \leq \|(I - P_K)\xi\|^2 \leq \xi ,$$

la dernière inégalité provenant du fait que $I - P_K$ est un projecteur orthogonal.

La deuxième assertion de la proposition résulte du fait que

$$\tilde{\theta} = \bar{\theta} + P_K(\hat{\theta} - \bar{\theta}) = (I - P_K)\,\bar{\theta} + P_K\hat{\theta}$$

et du fait que les projecteurs P_K et $I - P_K$ sont orthogonaux. □

3.1. Application aux fonctions splines de lissage. Nous supposons que la partie paramétrique du modèle des splines partielles est nulle et que la partie non paramétrique g de la moyenne du signal observé appartient à l'espace $\mathcal{H}^{(2m)}[0,1]$, mais sans forcément satisfaire aux conditions aux bords usuelles. Il n'est pas difficile de voir alors que l'on peut écrire g sous la forme suivante :

$$g(t) = \sum_{j=m}^{2m-1} [g^{(j)}(0)\ell_{j,0}(t) + g^{(j)}(1)\ell_{j,1}(t)] + g_0(t)$$

où $g_0^{(j)}(0) = g_0^{(j)}(1) = 0$, $j = m,\dots,2m-1$ et où les fonctions $\ell_{j,0}$ et $\ell_{j,1}$ sont des polynômes d'ordre 2m-1 déterminés par les conditions d'interpolation aux bords pour les dérivées et la condition $\int_0^1 \ell_{j,r}(t)dt = 0$, $r = 0,1$.

En notation vectorielle il vient :

$$\begin{pmatrix} g(t_1) \\ g(t_2) \\ \cdot \\ f(t_n) \end{pmatrix} = Q\gamma + \begin{pmatrix} g_0(t_1) \\ g_0(t_2) \\ \cdot \\ g_0(t_n) \end{pmatrix} \tag{3.2}$$

où Q est une matrice d'ordre n×2m déterminée par les valeurs des polynômes $\ell_{j,r}$ en les noeuds et les conditions d'interpolation aux bords.

Avec les notations du paragraphe 2 et en désignant par θ le vecteur de composantes les valeurs de g en les points t_i, posons

$$\hat{\theta} = Y$$

et

$$\bar{\theta} = S(\lambda)Y.$$

On voit alors que

$$E\{\bar{\theta}\} = S(\lambda)[Q\gamma + \theta_0] = \theta - (I - S(\lambda))Q\gamma - (I - S(\lambda))\theta_0 .$$

On peut donc appliquer la proposition 3.1. en prenant pour $\mathbf{K} = -(\mathbf{I} - S(\lambda))\mathbf{Q}$ et pour $\xi = -(\mathbf{I} - S(\lambda))\theta_0$. On obtient ainsi pour l'estimateur corrigé $\tilde{\theta}$ un biais majoré par $\|(I - S(\lambda))\theta_0\|^2$ qui est de l'ordre de λ^2 puisque θ_0 satisfait aux conditions aux bords. Le rang de $\mathbf{K}$ étant au plus égal à 2m et $\mu_{max}(\hat{\theta}) = \sigma^2$ on obtient donc pour notre estimateur corrigé le taux optimal du risque intégré $\mathcal{O}(n^{-\frac{4m}{4m+1}})$.

3.2. Le cas des splines partielles. Dans le cas général des splines inf-convolution la solution du problème de minimisation de la fonction de perte (1.3) de l'introduction est donnée par (cf. [3]) :

$$\hat{\beta} = ({}^t\mathbf{X}(I - S(\lambda))\mathbf{X})^{-1}\, {}^t\mathbf{X}(I - S(\lambda))\mathbf{Y},$$

et

$$\hat{f} = S(\lambda)(\mathbf{Y} - \mathbf{X}\hat{\beta}).$$

Rappelons alors la proposition suivante de Rice (cf. [7]) :

Proposition 3.2. En supposant que les instants d'observation sont régulièrement distribués, c'est-à-dire que $\int_0^{t_i} \omega(t)dt = \frac{2i-1}{2n}$, que $\frac{1}{n}{}^t\mathbf{X}\mathbf{X}$ converge vers une matrice définie positive quand $n \to \infty$, et que $\lambda \to 0$ et $n\lambda^{1/2m} \to \infty$ alors

$$\text{biais}(\hat{\beta}) = \mathcal{O}(\lambda^{1/2}) .$$

Si de plus $n\lambda^{1/m} \to \infty$, on a $\text{var}(\hat{\beta}) = \mathcal{O}(n^{-1})$. En supposant de plus que $n\lambda^2 \to 0$, on a

$$\text{biais}(\hat{f}_\lambda(t)) = \mathcal{O}(\lambda^{1/2})$$

et

$$\text{var}(\hat{f}_\lambda(t)) = \mathcal{O}(n^{-1}\lambda^{-1/2m}).$$

D'après cette proposition on voit que le biais et la variance de l'estimateur $\hat{f}_\lambda(t)$ de la partie non paramétrique du modèle conservent leurs taux de convergence que l'estimation de la partie paramétrique β soit entreprise ou non. Si l'on désire estimer de manière optimale (risque minimum) la partie lisse du signal, il faut donc prendre $\lambda \sim \mathcal{O}(n^{-\frac{2m}{2m+1}})$, comme dans le cas des splines de lissages classiques du paragraphe précédent. Mais avec cette vitesse de convergence de λ vers 0, on obtient

$$\text{biais}^2(\hat{\beta}) = \mathcal{O}(n^{-\frac{2m}{2m+1}}),$$

et

$$\text{var}(\hat{\beta}) = \mathcal{O}(n^{-1}).$$

Il s'ensuit que le carré du biais l'emporte sur la variance, rendant ainsi l'estimation de la composante paramétrique non consistante. Pour obtenir un estimateur de β convergent à un taux paramétrique, il faut a priori sous-lisser l'estimation de la composante non paramétrique. La proposition 3.1. nous permet encore une fois de corriger l'estimation sans être forcés de sous-lisser.

Plus précisément, notons encore une fois $S(\lambda)$ l'opérateur de lissage associé au problème des splines de lissages ($\beta = 0$) éventuellement corrigé pour les effets de bords selon la méthode du sous-paragraphe précédent. Posons

$$\bar{\theta} = S(\lambda)Y \quad \text{et} \quad \hat{\theta} = Y.$$

Le calcul donne

$$E\{\bar{\theta}\} = \theta - (I - S(\lambda))X\beta - (I - S(\lambda))f.$$

En prenant alors $K = -(I - S(\lambda))X$ et $\xi = -(I - S(\lambda))f$ dans la proposition, on obtient

$$\tilde{\theta} = X\tilde{\beta} + \tilde{f},$$

avec

$$\tilde{\beta} = ({}^tX^t(I - S(\lambda))(I - S(\lambda))X)^{-1}\, {}^tX^t(I - S(\lambda))(I - S(\lambda))Y$$

et

$$\tilde{f} = S(\lambda)(Y - X\tilde{\beta}).$$

Il est alors facile de voir que sous les mêmes conditions que la proposition 3.2 on a :

$$\text{biais}(\tilde{\beta}) = \mathcal{O}(\lambda) \quad \text{et} \quad \text{var}(\tilde{\beta}) = \mathcal{O}(n^{-1})$$

alors que

$$\text{biais}(\tilde{f}_\lambda) = \mathcal{O}(\lambda^{1/2}) \quad \text{et} \quad \text{var}(\tilde{f}_\lambda) = \mathcal{O}(n^{-1}\lambda^{-1/2m}) .$$

En prenant alors $\lambda = \mathcal{O}(n^{-\frac{2m}{2m+1}})$ on obtient

$$\text{biais}^2(\tilde{\beta}) = \mathcal{O}(\lambda^2) = \mathcal{O}(n^{-\frac{4m}{2m+1}}) < n^{-1} ,$$

ce qui permet l'estimation du paramètre linéaire sans sous lisser la partie non paramétrique.

Comme nous venons de le voir, le choix du paramètre λ de lissage est important. Les résultats précédents montrent, qu'au premier ordre, les propriétés asymptotiques de la fonction de risque intégré moyen restent essentiellement inchangées malgré la présence de la partie paramétrique. Ce résultat suggère que les méthodes de choix automatique du paramètre de lissage par validation croisée généralisée peuvent être utilisées en pratique pour l'estimation à taille d'échantillon fini.

Pour terminer notons ici que le modèle des splines partielles permet de traiter l'analyse de signaux bruités présentant un nombre fini de singularités en modèlisant par exemple le signal moyen sous la forme

$$f(t) = f_0(t) + \beta d_\alpha(t),$$

la fonction $d_\alpha(t) = \text{Ind}_{[\alpha,1]}$ représentant un saut d'ordre 0 en α. Le lecteur intéressé pourra consulter [1] pour une application de ce type de méthodes dans le cadre des processus ponctuels.

Remerciements. Je tiens à remercier le rapporteur pour sa lecture attentive. Ses remarques m'ont permis d'améliorer la présentation de ce travail.

REFERENCES

Antoniadis A. et Grégoire G. (1990) *Nonparametric estimation in change-points hazard models for censored survival data,* Ann. Stat. Soumis pour publication.

Rice J. et Rosenblatt M. (1981) *Integrated mean squared error of a smoothing spline,* Journal of Approximation Theory, n° 33, 353-369.

Girard D. et Laurent P.J. (1989) *Splines and estimation of nonlinear parameters,* Mathematical methods in CAGD, Academic Press, New York, pp. 273-298.

Eubank R.L. (1988) *Spline smoothing and nonparametric regression,* Marcell Dekker, Inc.

Heckman N.E. (1986) *Spline smoothing in a partly linear model,* J.R.S.S. serie B, n° 48, 244-248.

Laurent P.J. (1981) *Inf-convolution spline pour l'approximation de données discontinues* , RAIRO, Anal. Num., n° 20, 89-111.

Rice J. (1986) *Convergence rates for partially splined models,* Statist. and Prob. Letters, n° 4, 203-208.

Speckman P. (1985) *Spline smoothing and optimal rates of convergence in nonparametric regression models,* Ann. Statist., n° 13, 970-983.

Speckman, P. (1988) *Kernel smoothing in partially splined models,* JRSS B, 413-456.

Wahba G. (1984) *Partial spline models for the semi-parametric estimation of functions of several variables,* Statistical analysis of time series, I.M.S. Tokyo, pp. 319-329.